AD-A237 418

Systems
Optimization
Laboratory

DTIC
ELECTE
JUL 02 1991
S D
C

Primal-Dual Methods for Linear Programming

by
Philip E. Gill, Walter Murray,
Dulce B. Ponceleón and Michael A. Saunders

TECHNICAL REPORT SOL 91-3

May 1991

Access
NTIS G
DTIC

Department of Operations Research
Stanford University
Stanford, CA 94305

91-03882

91 6 01 051

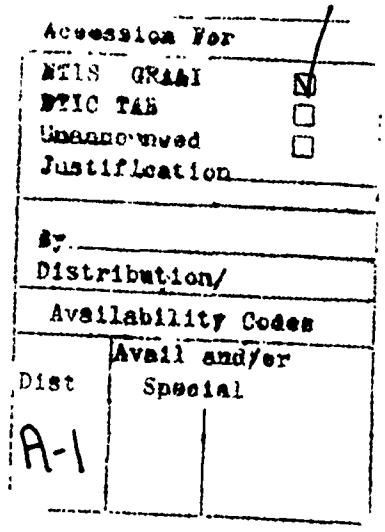
SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4022

Primal-Dual Methods for Linear Programming

by
Philip E. Gill, Walter Murray,
Dulce B. Ponceleón and Michael A. Saunders

TECHNICAL REPORT SOL 91-3

May 1991



Research and reproduction of this report were partially supported by the National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do NOT necessarily reflect the views of the above sponsors.

Reproduction in whole or in part is permitted for any purposes of the United States Government. This document has been approved for public release and sale; its distribution is unlimited.

PRIMAL-DUAL METHODS FOR LINEAR PROGRAMMING*

Philip E. GILL† Walter MURRAY‡
Dulce B. PONCELEÓN§ and Michael A. SAUNDERS‡

Technical Report SOL 91-3¶
May 1991

Abstract

Many interior-point methods for linear programming are based on the properties of the logarithmic barrier function. We first give a convergence proof for the (primal) projected Newton barrier method. We then analyze three types of barrier method that can be categorized as primal, dual and primal-dual. All three approaches may be derived from the application of Newton's method to different variants of the same system of nonlinear equations. A fourth variant of the same equations leads to a new primal-dual algorithm.

In each of the methods discussed, convergence is demonstrated without the need for a nondegeneracy assumption. In particular, convergence is established for a primal-dual algorithm that allows a different step in the primal and dual variables.

Finally, we describe a new method for treating free variables.

Keywords: linear programming, barrier methods, interior-point methods.

1. Introduction

For the most part we consider linear programs in the following standard form:

$$\begin{aligned} &\underset{x}{\text{minimize}} && c^T x \\ &\text{subject to} && Ax = b, \quad x \geq 0, \end{aligned} \tag{1.1}$$

where A is an $m \times n$ matrix with $m \leq n$. We focus on interior-point/barrier methods to solve (1.1).

---

*Presented at the Second Asilomar Workshop on Progress in Mathematical Programming, February, 1990, Asilomar, California.

†Department of Mathematics, University of California at San Diego, La Jolla, CA 92093, USA.

‡Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, CA 94305-4022, USA.

§Apple Computer, Inc., 20525 Mariani Avenue, Cupertino, CA 95014, USA.

¶The material contained in this report is based upon research supported by the National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

Initially we prove convergence for a *primal* barrier algorithm in which the iterates are assumed to satisfy $Ax = b$, but our main interest is in *primal-dual* algorithms that make as few assumptions as possible about the initial approximation to variables, and do not require a transformation of (1.1) into some mathematically equivalent linear program. We also allow liberal control of the barrier parameter.

A number of authors have described primal-dual algorithms that converge in polynomial time (see, e.g., Kojima, Mizuno and Yoshise [KMY89]; Monteiro and Adler [MA89]). However, such algorithms are generally *theoretical* and differ from the relatively few primal-dual algorithms that have been implemented for practical problems (see, e.g., McShane, Monma and Shanno [MMS89], Lustig, Marsten and Shanno [LMS89,LMS90], Mehrotra [Meh90], and Gill *et al.* [GMPS91]). Two key differences are the assumption that the step taken in the primal and dual spaces are the same and that the initial estimate of the solution is primal and dual feasible. It may be argued that the feasibility assumption is not overly restrictive because the linear program can be transformed into another problem with an identical solution, but a known feasible point. However, this ignores the possibility that the transformed problem may be more difficult to solve than the original.

Recently, Kojima, Megiddo and Mizuno [KMM90] have analyzed a primal-dual algorithm that is more similar to implemented algorithms. They define a steplength rule that allows (but does not guarantee) the possibility of different steps in the primal and dual spaces. They assume that the initial point is feasible.

The principal algorithms considered here do not require feasible iterates, and different steps may always be taken in the primal and dual spaces. These algorithms may be loosely categorized as primal, dual or primal-dual in order to distinguish between the different approaches. However, all of them are primal-dual in the sense that this term has been used for interior-point methods.

It is not within the scope of this paper to provide a numerical comparison of the different methods. Our intention is to give the methods a common setting and thereby highlight their similarities and differences. Our main purpose is to define and analyze *implementable* algorithms. For the purposes of analysis, it is necessary to include some procedures that are not present in standard implementations—the most notable of these being the definition of the steplength as the result of a linesearch instead of as a fixed fraction of the largest feasible step. However, the proposed linesearches are simple to implement and do not add significantly to the cost of an iteration. Moreover, the traditional "fixed" steplength usually satisfies the linesearch criteria. The proofs of convergence demonstrate that almost any step can be taken in the dual space. The existence of a wide range of steps for which convergence occurs may be the reason for the robustness of algorithms that do not incorporate a linesearch.

All the properties discussed apply to more general methods for problems in which some variables have upper bounds or are free. However, if the linear systems arising in the methods are solved using certain Schur complements, free variables become troublesome. In Section 8 we describe a new technique for the treatment of free variables.

The analysis presented here is applied only to linear programming. It has been

shown by Ponceleón [Pon90] that it may be generalized to indefinite quadratic programs.

1.1. Notation and Assumptions

Let x^* denote a solution to (1.1) and let $\mathcal{X}^*$ be the set of all solutions. Throughout we make the following assumptions:

(i) the constraint matrix A has full row rank;

(ii) the feasible region $S_0 = \{x \mid Ax = b,\ x \geq 0\}$ is compact; and

(iii) a *strictly* feasible point exists, i.e. there exists at least one point x such that $Ax = b$ and $x > 0$.

We shall use N to denote the matrix whose columns form a basis for the null space of A (thus $AN = 0$). Occasionally it will be necessary to refer to the i-th element of a sequence of vectors $\{x_i\}$ and the j-th component y_j of a vector y. To distinguish between x_i and y_j we shall use i to denote the i-th member of a sequence of vectors, and j to denote the j-th component of a vector. Unless otherwise stated, $\|\cdot\|$ refers to the vector two-norm or its induced matrix norm. The vector e denotes the vector $(1, 1, \ldots, 1)^T$.

2. Primal Barrier Methods

Barrier methods for linear programming generate approximations to both the primal and dual variables at each iteration. We shall use the term *primal method* to refer to a method that generates strictly positive values of the primal variables x, but does not restrict the values of the dual slack variables z. In the first algorithm we assume that the primal variables are feasible, i.e., that $Ax = b$. This assumption is relaxed for the remaining algorithms.

2.1. The Primal Barrier Subproblem

Barrier methods involve *major* and *minor* iterations. Each major iteration is associated with an element of a decreasing positive sequence of barrier parameters $\{\mu_k\}$ such that $\lim_{k\to\infty} \mu_k = 0$. The minor iterations correspond to an iterative process for the solution of the subproblem

$$\begin{array}{ll} \underset{x\in\Re^n}{\text{minimize}} & B(x,\mu) \equiv c^T x - \mu \sum_{j=1}^{n} \ln x_j \\ \text{subject to} & Ax = b, \end{array} \tag{2.1}$$

which is solved at every major iteration, i.e., for each value of $\mu = \mu_k$. Since $B(x,\mu)$ is a strictly convex function, there exists a unique minimizer $x^*(\mu)$ such that $Ax^*(\mu) = b$ and $x^*(\mu) > 0$.

Barrier methods are based on the fundamental result that $\lim_{\mu\to 0} x^*(\mu) \in \mathcal{X}^*$. For a proof of this result and a general discussion of barrier methods, see Fiacco and McCormick [FM68].

The special form of the derivatives of the barrier function makes Newton's method a natural choice for solving problem (2.1). At any given point x, Newton's method defines a search direction Δx such that $x + \Delta x$ continues to satisfy the linear constraints and minimizes a quadratic approximation to the barrier function. The vector Δx is the solution of the quadratic program

$$\begin{aligned} \underset{\Delta x}{\text{minimize}} \quad & \tfrac{1}{2}\Delta x^T H \Delta x + g^T \Delta x \\ \text{subject to} \quad & A\Delta x = 0, \end{aligned}$$

where $g(x,\mu) = c - \mu X^{-1}e$ and $H(x,\mu) = \mu X^{-2}$ are $\nabla B(x,\mu)$ and $\nabla^2 B(x,\mu)$, the gradient and Hessian of the barrier function, with $X = \text{diag}(x_j)$. If y denotes the Lagrange multiplier vector at x associated with the constraints $Ax = b$, the updated multipliers $y + \Delta y$ at $x + \Delta x$ satisfy

$$K \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = \begin{pmatrix} -g + A^T y \\ 0 \end{pmatrix}, \qquad \text{where} \qquad K \equiv \begin{pmatrix} H & A^T \\ A & \end{pmatrix}. \tag{2.2}$$

We shall refer to this system of equations as the *KKT system* and to the matrix K as the *KKT matrix*.

2.2. The Projected Newton Barrier Method

The formulation of the barrier subproblem (2.1) and the calculation of $x^*(\mu)$ by Newton's method was first embodied in the *projected Newton barrier method* of Gill *et al.* [GMSTW86]. The method requires the specification of two positive sequences: a bounded sequence $\{\delta_k\}$ that determines the accuracy of the solutions of (2.1) and a decreasing sequence of barrier parameters $\{\mu_k\}$ such that $\lim_{k\to\infty} \mu_k = 0$.

Algorithm PFP (Model Primal Feasible-Point Algorithm)

Compute x_0 such that $Ax_0 = b$, $x_0 > 0$;
Set $k = 0$, $i = 0$ and $i_k = 0$;
while not converged **do**
  Set $\mu = \mu_k$;
  while $\|N^T g(x_i,\mu)\| > \delta_k \mu$ **do**
    Find x_{i+1} such that
    $B(x_{i+1},\mu) < B(x_i,\mu)$, $x_{i+1} > 0$ and $Ax_{i+1} = b$;
    Set $i = i + 1$;
  end do;
  Set $k = k + 1$, $i_k = i$;
end do

Each member of the subsequence $\{x_{i_k}\}$ corresponds to an approximate minimizer of the subproblem (2.1) defined by μ_k. We shall refer to the consecutive indices of the sequence of minor iterations $i_{k-1}, i_{k-1} + 1, \ldots, i_k$ as $\mathcal{I}_k$.

Since $\lim_{k\to\infty} \mu_k = 0$, it follows that $\lim_{k\to\infty} \|x_{i_k} - x_k^*\| = 0$, where x_k^* is the nearest point to x_{i_k} in $\mathcal{X}^*$. The main difficulty lies in generating the sequence of

minor iterates $\{x_i,\ i \in \mathcal{I}_k\}$ so that the condition $\|N^T g(x_i,\mu)\| \leq \delta_k \mu_k$ is eventually satisfied. This issue is addressed in the next section.

The precise form of the termination condition for the minor iterations is not crucial. The only requirement is that the condition be satisfied in a neighborhood of $x^*(\mu_k)$ that shrinks to the point $x^*(\mu_k)$ as $k \to \infty$.

2.3. Convergence for the Primal Subproblem

In this section we show that the sequence $\{x_i,\ i = i_{k-1},\ldots\}$ generated by Newton's method with $\delta_k = 0$ converges to $x^*(\mu_k)$. It follows that for $\delta_k > 0$ the number of minor iterations required to satisfy the termination condition is finite.

Throughout this section we shall use the notation

$$\mu = \mu_k, \qquad B(x) = B(x,\mu), \qquad g(x) = g(x,\mu), \qquad H(x) = H(x,\mu),$$

to refer to quantities associated with the k-th subproblem.

The feasible set S_0 is compact by assumption. Given a positive constant θ and a feasible vector w such that $w \geq \theta e$, let $\Omega(w,\mu)$ denote the level set

$$\Omega(w,\mu) = \{x \mid B(x) \leq B(w)\}.$$

We have in mind w being the first minor iterate $x_{i_{k-1}}$ associated with μ and θ being the smallest component of w. Every subsequent minor iterate will lie in the set $S_0 \cap \Omega(w,\mu)$.

The essential element of the proof is the demonstration that the KKT matrix is bounded and has a bounded condition number at every point in the set

$$\hat{S} = S_0 \cap \Omega(w,\mu).$$

By assumption, A is bounded and has a bounded condition number. It follows that K will also have this property if H is bounded and has a bounded condition number. The latter properties in H follow from the following lemma, which shows that $\{(x_i)_j\}$ is bounded above and bounded away from zero.

Lemma 2.1. *Let θ be a positive constant, and let w be a given vector such that $w \geq \theta e$ and $Aw = b$. There exist positive constants σ_X and τ_X, independent of x, such that $\sigma_X e \leq x \leq \tau_X e$.*

Proof. The set $\hat{S}$ is compact since it is the intersection of the two closed sets S_0 and Ω, and it is a subset of the bounded set S_0. Since $\hat{S}$ is compact, there exists a constant τ_X such that $x_j \leq \tau_X$. The definition of $\hat{S}$ implies that every $x \in \hat{S}$ gives $B(x) \leq B(w)$. It follows that for all $x \in \hat{S}$,

$$c^T x - \mu \sum_{j=1}^{n} \ln x_j \leq c^T w - \mu \sum_{j=1}^{n} \ln w_j.$$

Therefore for each j,

$$-\mu \ln x_j \leq c^T w - \mu \sum_{j=1}^{n} \ln w_j - c^T x + \mu \sum_{r \neq j} \ln x_r.$$

Since $\hat{S}$ is compact, the quantities $\omega = \max\{|c^T x| \mid x \in \hat{S}\}$ and $\hat{\beta} = \max\{\ln x_j \mid x \in \hat{S}\}$ are bounded. Similarly, if $\theta > 0$, constant $\beta = \max\{\hat{\beta}, -\ln\theta\}$ is also bounded, and $-\mu \ln x_j \leq 2\omega + 2n\mu\beta$, or equivalently,

$$x_j \geq e^{-2(n\beta+\omega/\mu)} > 0,$$

as required. ∎

Corollary 2.1. *Let x be any element of $\hat{S}$. Let $H(x) = \mu X^{-2}$ where $X = \operatorname{diag}(x_j)$. Then there exist positive constants σ_H and τ_H, independent of x, such that for all vectors u,*

$$\sigma_H \|u\|^2 \leq u^T H(x) u \leq \tau_H \|u\|^2. \quad ∎$$

Lemma 2.2. *At every element of the sequence $\{x_i,\ i \in \mathcal{I}_k\}$ of Algorithm PFP, the matrix K is bounded and has a bounded condition number.* ∎

We now show that the sequence $\{x_i\}$ generated by Newton's method converges to $x^*(\mu)$, which implies that the condition $\|N^T g(x_i)\| \leq \delta_k \mu$ will be satisfied in a finite number of iterations.

The iterates of the projected Newton barrier method satisfy $x_{i+1} = x_i + \alpha_i \Delta x_i$, where the search direction Δx_i is defined by (2.2). The steplength α_i is determined from a linesearch, which locates a steplength that gives a sufficient decrease in $B(x)$. Throughout we shall use the Goldstein-Armijo conditions to define the steplength, although any of the standard steplength criteria would be suitable (see, e.g., Ortega and Rheinboldt [OR70]). For minimizing $B(x)$, the Goldstein-Armijo conditions are

$$\eta_1 \alpha_i \Delta x_i^T g(x_i) \leq B(x_i + \alpha_i \Delta x_i) - B(x_i) \leq \eta_2 \alpha_i \Delta x_i^T g(x_i), \tag{2.3}$$

where $0 < \eta_2 \leq \eta_1 < 1$.

Theorem 2.1. *Let $\{x_i\}$ be the sequence generated by Newton's method applied to the problem (2.1). Then $\lim_{i\to\infty} \|x_i - x^*(\mu)\| = 0$.*

Proof. Since a strictly feasible minimizer of the barrier function exists along Δx_i, there must exist a positive step α_i such that $x_i + \alpha_i \Delta x_i$ is a strictly feasible point and the Goldstein-Armijo conditions are satisfied. Consequently, $B(x_{i+1}) < B(x_i)$. Since $x_i \in \hat{S}$ and $\hat{S}$ is compact, it follows that $B(x_i)$ is bounded below and

$$\lim_{i\to\infty} \{B(x_{i+1}) - B(x_i)\} = 0. \tag{2.4}$$

Let $H(x)$ denote the Hessian matrix of $B(x)$. Since $x_i \in \hat{S}$, the corollary to Lemma 2.1 implies that there exists a positive constant σ_H such that

$$\Delta x_i^T H(x_i) \Delta x_i \geq \sigma_H \|\Delta x_i\|^2. \tag{2.5}$$

From (2.2) we have

$$\Delta x_i^T H(x_i) \Delta x_i = -\Delta x_i^T g(x_i).$$

Combining this identity with (2.5) and the second Goldstein-Armijo inequality (2.3) gives

$$B(x_i) - B(x_{i+1}) \geq -\eta_2 \alpha_i \Delta x_i^T g(x_i) \geq \eta_2 \sigma_H \alpha_i \|\Delta x_i\|^2,$$

which implies that $\lim_{i\to\infty} \alpha_i \|\Delta x_i\|^2 = 0$ from (2.4).

The Taylor-series expansion of $B(x_i + \alpha_i \Delta x_i)$ gives

$$B(x_{i+1}) = B(x_i) + \alpha_i \Delta x_i^T g(x_i) + \tfrac{1}{2}\alpha_i^2 \Delta x_i^T H(\hat{x}_i) \Delta x_i,$$

where $\hat{x}_i = x_i + \theta \alpha_i \Delta x_i$ for some $0 \leq \theta \leq 1$. Using this expansion in the first Goldstein-Armijo inequality (2.3) gives

$$\alpha_i \Delta x_i^T g(x_i) + \tfrac{1}{2}\alpha_i^2 \Delta x_i^T H(\hat{x}_i) \Delta x_i \geq \eta_1 \alpha_i \Delta x_i^T g(x_i),$$

and since $\Delta x_i^T g(x_i) < 0$, we have

$$|\Delta x_i^T g(x_i)| \leq \frac{\alpha_i}{2(1-\eta_1)} \Delta x_i^T H(\hat{x}_i) \Delta x_i. \tag{2.6}$$

Since $\hat{S}$ is convex, $\hat{x}_i \in \hat{S}$ and it follows from the corollary to Lemma 2.1 that there exists a constant τ_H such that

$$\Delta x_i^T H(\hat{x}_i) \Delta x_i \leq \tau_H \|\Delta x_i\|^2.$$

Combining this inequality with (2.6) gives

$$|\Delta x_i^T g(x_i)| \leq \frac{\alpha_i}{2(1-\eta_1)} \Delta x_i^T H(\hat{x}_i) \Delta x_i \leq \frac{\tau_H}{2(1-\eta_1)} \alpha_i \|\Delta x_i\|^2.$$

Since $\lim_{i\to\infty} \alpha_i \|\Delta x_i\|^2 = 0$, we obtain

$$\lim_{i\to\infty} \Delta x_i^T g(x_i) = 0. \tag{2.7}$$

From (2.2) we have

$$N^T H(x_i) N \Delta x_{N_i} = -N^T g(x_i), \tag{2.8}$$

where $\Delta x_i = N \Delta x_{N_i}$. Since $N^T H(x_i) N$ is bounded and has a bounded condition number, it follows from (2.7) and (2.8) that $\lim_{i\to\infty} \Delta x_i = 0$ and $\lim_{i\to\infty} N^T g(x_i) = 0$. Since $x^*(\mu)$ is the unique feasible point for which $N^T g(x^*(\mu)) = 0$, we have $\lim_{i\to\infty} \|x_i - x^*(\mu)\| = 0$ as required. ∎

3. Getting Feasible

There are various ways to eliminate the requirement of Algorithm PFP that an initial strictly feasible point be known.

3.1. An Artificial Variable

A common approach is to introduce an additional variable, or set of variables, and minimize a composite objective function. For example, given $x_0 > 0$, consider the transformed problem

$$\begin{aligned} &\underset{x\in\Re^n,\ \xi\in\Re}{\text{minimize}} && c^Tx + \rho\xi \\ &\text{subject to} && Ax + \xi u = b, \quad x \geq 0, \quad \xi \geq -1, \end{aligned}$$

where $u = (b - Ax_0)/\|b - Ax_0\|$ and ρ is a positive scalar. The initial value of ξ is $\|b - Ax_0\|$, so a strictly feasible point for the transformed problem is known. If a step would make ξ negative during an iteration, a shorter step is taken to make $\xi = 0$. Once ξ is zero, it is eliminated from the problem.

The difficulty with this and similar approaches lies in choosing the value for the parameter ρ. Although ρ must be sufficiently large, if it is chosen too large, the infeasibilities dominate the objective function and the method behaves like a two-phase algorithm. If no strictly feasible point exists, the efficiency of the algorithm can depend critically on the choice of ρ.

3.2. A Merit Function

The method of Section 2.1 may be generalized so that Δx is the solution of the quadratic program

$$\begin{aligned} &\underset{\Delta x}{\text{minimize}} && \tfrac{1}{2}\Delta x^T H \Delta x + g^T\Delta x \\ &\text{subject to} && A\Delta x = b - Ax \end{aligned}$$

and satisfies

$$K\begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = \begin{pmatrix} -g + A^Ty \\ b - Ax \end{pmatrix}. \tag{3.1}$$

We may then introduce a *merit function* that balances the aims of minimizing $B(x,\mu)$ and reducing some norm of $Ax - b$. For example, one possible merit function is

$$M(x,\rho) = B(x,\mu) + \rho\|Ax - b\|_1,$$

where ρ is chosen suitably large. It can be shown that if Δx is defined by (3.1) then it is a descent direction for $M(x,\rho)$ (see Section 7). We may prove convergence in a manner similar to that for the feasible-point algorithm. It may be thought that this approach also depends on choosing a "good" value for the parameter ρ. However, ρ affects only the steplength and not the *direction* of search. Moreover, it is relatively trivial to adjust ρ dynamically. We can take the step we would like to take and then check whether a suitable value of ρ exists for which the linesearch criteria are satisfied.

We shall not pursue this approach with respect to primal barrier algorithms, since we think a better approach is outlined in the next section. However, we shall return to this merit function when we discuss a primal-dual method in Section 7.

3.3. Newton's Method Applied to the Optimality Conditions

Since $B(x,\mu)$ is strictly convex, $x^*(\mu)$ is the only strictly feasible constrained stationary point for problem (2.1). Therefore, an alternative method for finding $x^*(\mu)$ is to use Newton's method for *nonlinear equations* to find the stationary point of the Lagrangian function $L(x,y) = B(x,\mu) - y^T(Ax-b)$. Since the gradient of $L(x,y)$ is zero at $x^*(\mu)$, we obtain the $n+m$ nonlinear equations

$$\nabla L(x,y) = \begin{pmatrix} c - \mu X^{-1}e - A^T y \\ Ax - b \end{pmatrix} = 0, \tag{3.2}$$

whose Jacobian is closely related to the KKT matrix K. The solution of the KKT system (3.1) is a descent direction for $\|\nabla L\|$, and a steplength may be chosen to achieve a sufficient reduction in $\|\nabla L\|$. As in Algorithm PFP, this merit function ensures that x_j cannot be arbitrarily close to its bound.

We now extend this approach to obtain the principal algorithms of interest in this paper.

4. A Primal Primal-Dual Method

Following common practice, we introduce a third vector of variables $z = c - A^T y$. We now wish to solve the nonlinear equations $f_P(z,x,y) = 0$, where

$$f_P(z,x,y) \equiv \begin{pmatrix} \bar{f} \\ \hat{f} \\ r \end{pmatrix} = \begin{pmatrix} z - \mu X^{-1}e \\ c - A^T y - z \\ Ax - b \end{pmatrix}. \tag{4.1}$$

When it is necessary to consider the full vector of variables z, x and y, the vector v will denote the $(2n+m)$-vector $(z,\ x,\ -y)$. The symbols $f_P(z,x,y)$ and $f_P(v)$ will be used interchangeably for f_P, depending on the point of emphasis. The Newton direction $\Delta v = (\Delta z,\ \Delta x,\ -\Delta y)$ satisfies the linear system

$$J_P \Delta v = -f_P, \qquad \text{where} \qquad J_P = \begin{pmatrix} I & \mu X^{-2} & 0 \\ -I & 0 & A^T \\ 0 & A & 0 \end{pmatrix}. \tag{4.2}$$

Apart from the last block of columns being multiplied by -1, J_P is the Jacobian of the nonlinear equations (4.1). We shall refer to J_P as the Jacobian.

The directions Δx and Δy from (4.2) are identical to those defined by the KKT system (3.1), and to those associated with (3.2). However, for the nonlinear equations $\nabla L(x,y) = 0$ and $f_P(z,x,y) = 0$, the steplength is chosen to produce a sufficient decrease in $\|\nabla L\|^2$ and $\|f_P\|^2$ respectively. In the latter case, the Goldstein-Armijo conditions give the following conditions on α_i:

$$0 < -2\eta_2\alpha_i \Delta v^T J_P^T f_P(v_i) \le \|f_P(v_i)\|^2 - \|f_P(v_{i+1})\|^2 \le -2\eta_1\alpha_i \Delta v^T J_P^T f_P(v_i).$$

Since $J_P \Delta v = -f_P(v_i)$, these conditions can be restated in the equivalent form

$$0 < 2\eta_2\alpha_i \leq 1 - \frac{\|f_P(v_{i+1})\|^2}{\|f_P(v_i)\|^2} \leq 2\eta_1\alpha_i,$$

which is easily tested.

Since the residuals $\hat{f}$ and r are linear in x, y and z, they are simply related to their values in the previous iteration. Suppose that r and $\hat{f}$ are nonzero at iteration i. After a step of Newton's method with steplength α_i, we have

$$r_{i+1} = (1 - \alpha_i)r_i \quad \text{and} \quad \hat{f}_{i+1} = (1 - \alpha_i)\hat{f}_i. \tag{4.3}$$

At the first iteration $\|z_0\|$ and $\|y_0\|$ are bounded and x_0 is bounded away from zero, which implies that the Jacobian is bounded and has a bounded condition number. It follows that $\alpha_0 > 0$. Hence the relations (4.3) imply that $r_i = \gamma_i r_0$ for some scalar γ_i such that $0 \leq \gamma_i < 1$. If a unit step is taken at any iteration, $\hat{f}$ and r will be zero for all subsequent iterations.

The complete algorithm is as follows.

Algorithm PPD (Model Primal Primal-Dual Algorithm)

Set v_0, with $x_0 > 0$ and $z_0 > 0$;
Set $k = 0$, $i = 0$ and $i_k = 0$;
while not converged **do**
  Set $\mu = \mu_k$;
  while $\|f_P(v_i, \mu)\| > \delta_k \mu$ **do**
    Find v_{i+1} such that
    $\|f_P(v_{i+1}, \mu)\|^2 < \|f_P(v_i, \mu)\|^2$ and $x_{i+1} > 0$;
    Set $i = i + 1$;
  end do;
  Set $k = k + 1$, $i_k = i$;
end do

4.1. Convergence

The convergence proof for this algorithm is similar to that for Algorithm PFP in that it is necessary to show that for each barrier subproblem, J_P remains bounded and has a bounded condition number. However, in Algorithm PFP the iterates lie in S_0, whereas here it is not obvious that the iterates $\{x_i\}$ lie in any compact set. We establish this fact in the next lemma and then show that $\{v_i\}$ lies in a compact set.

Lemma 4.1. *Let τ_r denote a positive constant. If the feasible region S_0 is compact, then so is the set*

$$S_A = \{x \mid x \geq 0, \ \|Ax - b\| \leq \tau_r\}.$$

Proof. Since S_A is closed, it only remains to be shown that S_A is bounded. Since the elements of w are nonnegative, it follows that S_A will be compact if $e^T w$ is bounded.

Let w be any member of S_A and let r_w denote the residual $Aw - b$ associated with w. Let w^* be a solution of the linear program

$$\begin{aligned} &\underset{w \in \mathbb{R}^n}{\text{maximize}} \quad e^T w \\ &\text{subject to} \quad Aw = b + r_w, \quad w \geq 0. \end{aligned} \tag{4.4}$$

It follows that S_A is compact if $e^T w^*$ is bounded.

Let R denote a full-rank matrix whose columns form a basis for the range space of A^T. It follows that w may be expressed as

$$w = Nw_N + Rw_R. \tag{4.5}$$

In particular, $w^* = Nw_N^* + Rw_R^*$, and substitution in (4.4) gives

$$ARw_R^* = b + r_w.$$

Since $\|r_w\| \leq \tau_r$ and AR has full column rank, this equation implies that $\|w_R^*\|$ is bounded. Equation (4.5) now implies that w_N^* is a solution of

$$\begin{aligned} &\underset{w_N}{\text{maximize}} \quad e^T w_N \\ &\text{subject to} \quad Nw_N \geq -Rw_R^*. \end{aligned} \tag{4.6}$$

Assume that the linear program (4.6) is unbounded. Then there must exist a nontrivial feasible direction u such that $Nu \geq 0$. If x is any point in S_0, then $x + \gamma Nu$ must also be in S_0 for any positive γ, which contradicts the compactness of S_0. Consequently, the solution of (4.6) is bounded and S_A is compact. ∎

Lemma 4.2. *Let r_0 denote the residual $r_0 = Ax_0 - b$, with $x_0 > 0$. Define the set*

$$S_r = \{(z,x,y) \mid x \geq 0,\ Ax - b = \gamma r_0\}$$

for some γ, $0 \leq \gamma < 1$, and the level set

$$\Gamma(\tau_f, \mu) = \{(z,x,y) \mid \|f_P(z,x,y)\| \leq \tau_f\}.$$

Then the set $\hat{S} = S_r \cap \Gamma(\tau_f, \mu)$ is compact.

Proof. Throughout this proof we shall assume that v is a vector in $\hat{S}$. From the definition of S_r, we have $\|Ax - b\| \leq \|r_0\|$ and it follows from Lemma 4.1 that the x-components of v are bounded. It remains to be shown that the y and z components of v are bounded. Note that the components of both $\bar{f}$ and $\hat{f}$ are bounded since they are components of the bounded vector f_P.

Consider the equations $\bar{f} = z - \mu X^{-1}e$ of (4.1). Premultiplying $\bar{f}$ by x^T and using the fact that both x and $\bar{f}$ are bounded, we see that there exists a constant τ_1 such that

$$x^T z = x^T \bar{f} + \mu n < \tau_1. \tag{4.7}$$

Also, since $x \geq 0$,

$$z_j > \bar{f}_j > -\tau_2 \tag{4.8}$$

for some positive constant τ_2.

If x^T is now applied to the second equation of (4.1), $\hat{f} = c - A^T y - z$, we obtain

$$x^T \hat{f} = x^T c - x^T A^T y - x^T z = x^T c - (b^T + r^T)y - x^T z.$$

Simple rearrangement gives

$$-(b^T + r^T)y = x^T \hat{f} + x^T z - x^T c, \tag{4.9}$$

and it follows from (4.7) and the bounds on $\hat{f}$ and x that

$$-(b^T + r^T)y < \tau_3. \tag{4.10}$$

Similarly, using $x = x_0$ in (4.9) gives

$$(b^T + r_0^T)y = -x_0^T \hat{f} - x_0^T z + x_0^T c \leq -x_0^T \hat{f} - \sum_{J_-} (x_0)_j z_j + x_0^T c,$$

where J_- is the set of indices of the *negative* elements of z (recall that the elements of x_0 are positive). It follows from (4.8) that

$$(b^T + r_0^T)y < \tau_4. \tag{4.11}$$

Using (4.10) and the assumption that $r = \gamma r_0$ for some $0 \leq \gamma < 1$ gives

$$-(b^T + \gamma r_0^T)y < \tau_3. \tag{4.12}$$

Combining (4.11) and (4.12) gives

$$-b^T y \leq \frac{\tau_3 + \gamma \tau_4}{1 - \gamma} < \tau_5.$$

It now remains to bound the term $x^{*T} z$. Using (4.9) with $x = x^*$ gives

$$x^{*T} z = x^{*T} c - x^{*T} \hat{f} - b^T y.$$

Since $x_j^* > 0$ and $\|x^*\|$ is bounded (see Lemma 2.1), all the terms on the right-hand side of this expression are bounded, with $x^{*T} z < \tau_6$ for some positive constant τ_6. Lemma 2.1 also implies the existence of positive constants σ_X and τ_X such that $\sigma_X \leq x_j^* \leq \tau_X$. Clearly $\|z_j\|$ is bounded, with

$$z_j < (\tau_6 + n\tau_X \tau_2)/\sigma_X.$$

Since A has full row rank, the bounds on $\|\hat{f}\|$ and $\|z\|$ in the equation $\hat{f} = c - A^T y - z$ imply that $\|y\|$ is bounded, as required. ∎

Lemma 4.3. *If $v \in \hat{S}$ then J_P is bounded and has a bounded condition number.*

Proof. It is enough to show that x_j is bounded away from zero if $v \in \hat{S}$. We have from (4.1) that $z_j - \bar{f}_j = \mu/x_j$. Hence

$$|z_j| + \|\bar{f}\| \geq \mu/x_j \quad \text{or equivalently} \quad x_j \geq \frac{\mu}{|z_j| + \|\bar{f}\|}.$$

It follows from Lemma 4.2 that there exists a positive constant τ_z such that $|z_j| < \tau_z$ for all $v \in \hat{S}$, and by assumption, $\|\bar{f}\| \leq \tau_f$. Hence, $x_j \geq \mu/(\tau_z + \tau_f) > 0$.

From Lemma 4.1, x is uniformly bounded above. Since x_j is bounded away from zero, J_P is bounded and the condition number of J_P is bounded. ∎

The proof of the following theorem is similar to that for Theorem 2.1.

Theorem 4.1. *Let $\{v_i\}$ be the sequence generated by Newton's method applied to the equations (4.1). Then* $\lim_{i\to\infty} \|v_i - v^*(\mu)\| = 0$. ∎

It follows that Newton's method generates a point that satisfies the condition $\|f_P(v_i, \mu)\| \leq \delta_k \mu$ in a finite number of iterations.

5. Summary of Primal Methods

In all the algorithms considered so far (excluding the artificial-variable method of Section 3.1), the search directions for x and y are the same as those given by (4.2). The steplength α may be chosen to reduce one of the following functions:

(i) $M(x, \rho) = B(x, \mu) + \rho\|Ax - b\|_1$ (search in x-space).

(ii) $\|c - \mu X^{-1}e - A^T y\|^2 + \|Ax - b\|^2$ (search in x and y-space).

(iii) $\|c - z - A^T y\|^2 + \|z - \mu X^{-1}e\|^2 + \|Ax - b\|^2$ (search in x, y and z-space).

The only additional restriction on α is the requirement that $x + \alpha\Delta x > 0$. In all cases, approximations in the x, y and z-space may be generated even though they are *necessary* only in (iii). Thus, all three methods may be viewed as primal-dual algorithms.

If some steplength other than α is taken along Δz and Δy, a sequence of *auxiliary* y and z-values can be generated that approximate y^* and z^*. For this sequence, a different step α_z in the y and z-space is needed to maintain $z > 0$. Since α_z is not usually equal to α, a dual feasible point may be found before a primal feasible point (or vice versa). Provided that the step taken in the y-space is also α_z, once a dual feasible point is found, all subsequent approximations will be dual feasible.

One advantage of (ii) and (iii) is that it is not necessary to compute logarithms. Moreover, it is not necessary to define a parameter ρ that balances feasibility and optimality, although it may be advantageous to weight the norms occurring in (ii) and (iii).

6. Dual Methods

The dual of the linear program (1.1) may be written as

$$\begin{aligned} &\underset{y,z}{\text{minimize}} && -b^T y \\ &\text{subject to} && c - A^T y - z = 0, \quad z \geq 0. \end{aligned} \tag{6.1}$$

The dual barrier subproblem is

$$\begin{aligned} &\underset{y \in \Re^m,\, z \in \Re^n}{\text{minimize}} && -b^T y - \mu \sum_{j=1}^{n} \ln z_j \\ &\text{subject to} && c - A^T y - z = 0. \end{aligned} \tag{6.2}$$

Newton's method applied to this problem defines the y-space search direction from a system similar to (2.2). (The right-hand side is changed and $H = (1/\mu)Z^2$, where $Z = \text{diag}(z_j)$.) Given an initial feasible point (y_0, z_0) we may define a dual algorithm DFP analogous to PFP.

Similarly, we may construct an algorithm based upon the optimality conditions for (6.2):

$$\begin{aligned} x - \mu Z^{-1} e &= 0, \\ c - A^T y - z &= 0, \\ Ax - b &= 0. \end{aligned} \tag{6.3}$$

As noted by Megiddo [Meg89], the solution of these equations is identical to the solution of (4.1). Newton's method applied to (6.3) solves the linear system $J_D \Delta v = -f_D$, where

$$f_D(z,x,y) \equiv \begin{pmatrix} \bar{f} \\ \hat{f} \\ r \end{pmatrix} = \begin{pmatrix} x - \mu Z^{-1} e \\ c - A^T y - z \\ Ax - b \end{pmatrix} \quad \text{and} \quad J_D = \begin{pmatrix} \mu Z^{-2} & I & 0 \\ -I & 0 & A^T \\ 0 & A & 0 \end{pmatrix}.$$

The resulting algorithm, DPD, is identical to PPD except that J_P and f_P are replaced by J_D and f_D, and the z-variables are restricted during the linesearch instead of the x-variables. It can be shown that like J_P, the matrix J_D remains bounded and has a bounded condition number. Moreover, a step satisfying the Goldstein-Armijo conditions must exist, since $\|f_D\|$ would be infinitely large if any element of z were zero. Note that whenever every component of z is positive, Δv is a descent direction for $\|f_D\|^2$.

As in algorithm PPD, an auxiliary sequence can be generated by allowing the the primal and dual steplengths to be different. In this case, the sequence would be a strictly positive approximation to x^*.

Theorem 6.1. *Let $\{v_i\}$ be the sequence generated by Newton's method applied to the equations (6.3). Then $\lim_{i \to \infty} \|v_i - v^*(\mu)\| = 0$.*

Proof. Let $v = (z, x, -y)$. It follows immediately from $z > 0$ and

$$-\mu Z^{-1}e + x = \bar{f}$$

that $x_i \geq -\tau_1 > -\infty$ and Lemma 4.1 implies that x lies in a compact set. An identical argument to that used in Lemma 4.2 shows that v also lies in a compact set. It follows immediately that $\|Z^{-1}\|$ is bounded. Hence J_D is bounded and has a bounded condition number. The required result follows from an identical argument to that of Theorem 2.1. ∎

7. Primal-Dual Algorithms

7.1. A Primal-Dual Method

Algorithms PPD and DPD both generate a sequence of approximations to $v^*(\mu)$. In addition, $v^*(\mu)$ solves the system of equations $f_{PD}(z, x, y) = 0$, where

$$f_{PD}(z, x, y) \equiv \begin{pmatrix} \bar{f} \\ \hat{f} \\ r \end{pmatrix} = \begin{pmatrix} Xz - \mu e \\ c - A^T y - z \\ Ax - b \end{pmatrix}. \tag{7.1}$$

Newton's method for these nonlinear equations gives the linear system

$$J_{PD}\Delta v = -f_{PD}, \qquad \text{where} \qquad J_{PD} = \begin{pmatrix} X & Z & 0 \\ -I & 0 & A^T \\ 0 & A & 0 \end{pmatrix}, \tag{7.2}$$

which has been used by Lustig, Marsten and Shanno [LMS89,LMS90], Mehrotra [Meh90], and Gill *et al.* [GMPS91] (see also Lustig [Lus88]). Methods based on the solution of (7.2) are usually referred to as *primal-dual* algorithms because both x and z are maintained to be positive. It must be stressed that this terminology does not imply any direct connection between (7.2) and the primal-dual form of LP. If the latter is transformed using a barrier function, the resulting optimality conditions involve six sets of variables and two independent systems of equations that are identical to (4.1) and (6.3).

Unlike J_P and J_D, J_{PD} is independent of μ. If α is chosen to maintain sufficient positivity in both x and z, J_{PD} will be a bounded matrix with a bounded condition number. A key difference with these equations is that it is no longer obvious that if α is chosen to satisfy the Goldstein-Armijo conditions then a suitable step that maintains both $z > 0$ and $x > 0$ exists. We therefore propose an algorithm that takes a different step in the x-space than in the y and z-space and uses $M(x, \rho)$ as a merit function rather than $\|f_P\|^2$. If σ_Z, τ_Y and τ_Z are preassigned positive constants, let S_Y and S_Z be the sets

$$S_Y = \{y \mid \|y\| \leq \tau_Y\} \quad \text{and} \quad S_Z = \{z \mid 0 < \sigma_Z e \leq z \leq \tau_Z e\}.$$

Algorithm PD (Model Primal-Dual Algorithm)

Set v_0, with $x_0 > 0$, $z_0 \in S_Z$ and $y_0 \in S_Y$;
Set $k = 0$, $i = 0$ and $i_k = 0$;
while not converged **do**
  Set $\mu = \mu_k$;
  while $\|N^T g(x_i, \mu)\| + \|r\| > \delta_k \mu$ **do**
    Select any $z_{i+1} \in S_Z$ and $y_{i+1} \in S_Y$;
    Solve $J_{PD} \Delta v_i = -f_{PD}$ for Δx_i;
    Find $x_{i+1} = x_i + \alpha_i \Delta x_i$ such that
    $M(x_{i+1}, \rho) < M(x_i, \rho)$ and $x_{i+1} > 0$;
    Set $i = i + 1$;
  end do;
  Set $k = k + 1$, $i_k = i$;
end do

The convergence of Algorithm PD follows directly if it can be shown that (7.2) generates a sequence $\{x_i\}$ converging to $x^*(\mu)$.

Given positive constants τ_r and τ_M, define the level set

$$\bar{S} = \{x \mid \|Ax - b\| \leq \tau_r, \ M(x, \rho) \leq \tau_M\}.$$

Similar arguments used to prove Lemma 4.1 show that $\bar{S}$ is compact.

Lemma 7.1. *If $x \in \bar{S}$ then there exists a positive σ_X, independent of x, such that $x_i \geq \sigma_X e$.* ∎

Lemma 7.2. *Given positive constants τ_r, τ_Y, τ_X and τ_Z, assume that x, y and z satisfy $\|r\| = \|Ax - b\| \leq \tau_r$, $\|y\| < \tau_Y$, $0 < x < \tau_X e$ and $0 < z < \tau_Z e$. Then there exist constants ρ, γ ($\gamma > 0$) and β ($\beta \geq 1$) such that*

$$\Delta x^T \nabla M(x, \rho) \leq -\gamma \|N^T g\|^2 - \beta \|r\|_1,$$

where Δx is defined by (7.2).

Proof. If the system

$$\begin{pmatrix} ZX^{-1} & A^T \\ A & \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = - \begin{pmatrix} c - \mu X^{-1} e - A^T y \\ Ax - b \end{pmatrix} \tag{7.3}$$

is solved for Δx and Δy, it follows from the assumptions that $\|\Delta x\|$ is bounded.

Observe that the right-hand side of (7.3) is identical to that of (2.2). It follows that

$$H_N \Delta x_N = -N^T (g + ZX^{-1} A^T \Delta x_A), \tag{7.4}$$

where $g = \nabla B(x, \mu)$, $H_N = N^T Z X^{-1} N$, $\Delta x = N \Delta x_N + A^T \Delta x_A$, and

$$A A^T \Delta x_A = -r. \tag{7.5}$$

Consider now the merit function $M(x,\rho)$ given in Section 3. By definition we have

$$\Delta x^T \nabla M(x,\rho) = \Delta x_N^T N^T (g + \rho A^T \bar{e}) + \Delta x_A^T A(g + \rho A^T \bar{e}),$$

where $\bar{e}$ is a vector whose elements are one in magnitude and whose signs are the same as r. Define $u \equiv (AA^T)^{-1}A(I - X^{-1}ZNH_N^{-1}N^T)g$. Substituting for Δx_N from (7.4) and Δx_A from (7.5) gives

$$\begin{aligned}\Delta x^T \nabla M(x,\rho) &= -g^T N H_N^{-1} N^T g - r^T u - \rho r^T \bar{e}, \\ &\leq -\gamma \|N^T g\|^2 - \beta \|r\|_1,\end{aligned}$$

where $\gamma > 0$ is the reciprocal of the largest eigenvalue of H_N, $\beta \geq 1$, and ρ is chosen such that

$$\rho = \max(1 - \frac{r^T u_M}{r^T \bar{e}},\ 0),$$

with u_M the vector u evaluated at the point $x \in \bar{S}$ for which $r^T u$ has its minimum value. ∎

Lemma 7.3. *If Δx is defined by (7.2), there exist positive α and σ_X such that the Goldstein-Armijo conditions are satisfied with $x + \alpha \Delta x > \sigma_X e$.*

Proof. If α_M is the largest feasible step along Δx, then $M(x + \alpha_M \Delta x, \rho)$ is infinite and it follows that there exists a positive number α^* that solves the problem of $\min_\alpha\ M(x + \alpha \Delta x, \rho)$ subject to $x + \alpha \Delta x > 0$. Hence, a strictly feasible point exists for which the Goldstein-Armijo conditions are satisfied. ∎

Lemma 7.4. *Let σ_Z, τ_Y and τ_Z be preassigned positive constants. Consider sequences $\{z_i\}$ and $\{y_i\}$ such that $\sigma_Z e \leq z_i \leq \tau_Z e$ and $\|y_i\| \leq \tau_Y$. Let $\{x_i\}$ denote the sequence $x_{i+1} = x_i + \alpha_i \Delta x_i$ and $x_0 > 0$, where Δx_i is defined by (7.1) and α_i satisfies the Goldstein-Armijo conditions on $M(x,\rho)$ with the requirement that $x_{i+1} > 0$. If ρ is sufficiently large (but bounded) then $\lim_{i\to\infty} x_i = x^*(\mu)$.*

Proof. Since $\{x_i\}$ lies in a compact set, it follows that x_i is bounded for all i. Moreover, since x_i lies in $\bar{S}$, there exists a positive σ_X such that $x_i \geq \sigma_X e$ for all i. Every element of the sequence $\{x_i\}$ satisfies the assumptions of Lemma 7.2 and we have

$$\Delta x_i^T \nabla M(x_i,\rho) \leq -\gamma \|N^T g(x_i)\|^2 - \beta \|r(x_i)\|_1,$$

where $\gamma > 0$ and $\beta \geq 1$. It follows from Lemma 7.2 that $\{M(x_i,\rho)\}$ is a strictly monotonically decreasing sequence. Since $\{x_i\} \in \bar{S}$, it follows that $\{M(x_i,\rho)\}$ must converge and the Goldstein-Armijo conditions give

$$\lim_{i\to\infty} \alpha_i \Delta x_i^T \nabla M(x_i,\rho) = \lim_{i\to\infty} \alpha_i (\gamma \|N^T g(x_i)\|^2 + \beta \|r(x_i)\|_1) = 0.$$

The proof now follows a similar argument to that given in Lemma 2.1. ∎

Lemma 7.5. *If the assumptions and definitions of Lemma 7.4 hold then*

$$\lim_{i\to\infty} y_i + \Delta y_i = y^*(\mu) \quad \textit{and} \quad \lim_{i\to\infty} z_i + \Delta z_i = z^*(\mu).$$

Proof. It follows from (7.3) and the optimality conditions of (2.1) that

$$\lim_{i\to\infty} y_i + \Delta y_i = y^*(\mu) \quad \text{and} \quad \lim_{i\to\infty} \Delta x_i = 0.$$

From (7.2) we have $Z_i(x_i + \Delta x_i) = -X_i \Delta z_i + \mu e$. Since $\lim_{i\to\infty} x_i = x^*(\mu)$ and $\lim_{i\to\infty} \Delta x_i = 0$, we have

$$\lim_{i\to\infty} z_i + \Delta z_i = \mu X_i^*(\mu)^{-1} e = z^*(\mu),$$

where $X_i^*(\mu) = \text{diag}((x_i^*(\mu))_j)$. ∎

The above result shows that even for quite arbitrary choices of $\{z_i\}$ and $\{y_i\}$, approximations to $y^*(\mu)$ and $z^*(\mu)$ may be obtained.

In practice, certain choices of z_i and y_i lead to more efficient algorithms. The primal algorithm of Section 3.2 using the merit function $M(x,\rho)$ may be viewed as being equivalent to Algorithm PD with z_i chosen as $\mu X_i^{-1} e$ and $y_{i+1} = y_i + \alpha_i \Delta y_i$. Since $\|Ax - b\|_1$ is implicitly bounded by the linesearch, Lemma 4.1 implies that x_i is bounded. It follows that each $(z_i)_j$ is bounded away from zero, and $z_i \in S_Z$ for suitably small σ_Z.

Alternatively, values of y and z may be determined from a linesearch. A steplength θ_i in the z and y-space can be taken as an approximate solution of the univariate problem

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \|f_{PD}(z_i + \theta \Delta z_i, x_i, y_i + \theta \Delta y_i)\|^2 \\ \text{subject to} \quad & z_i + \theta \Delta z_i \geq \eta \mu X_{i+1}^{-1} e, \qquad \theta \leq 1, \end{aligned}$$

where η is some preassigned constant in $(0,1]$.

7.2. Another Primal-Dual Algorithm

A second primal-dual algorithm can be derived by observing that $v^*(\mu)$ solves the system of equations $f_{PDD}(z,x,y) = 0$, where

$$f_{PDD}(z,x,y) \equiv \begin{pmatrix} \bar{f} \\ \hat{f} \\ r \end{pmatrix} = \begin{pmatrix} \mu X^{-1} Z^{-1} e - e \\ c - A^T y - z \\ Ax - b \end{pmatrix}. \tag{7.6}$$

Newton's method for these equations gives the linear system $J_{PDD} \Delta v = -f_{PDD}$, where

$$J_{PDD} = \begin{pmatrix} -\mu Z^{-2} X^{-1} & -\mu X^{-2} Z^{-1} & 0 \\ -I & 0 & A^T \\ 0 & A & 0 \end{pmatrix}.$$

Unlike the primal-dual method of Section 7.1, there always exists a steplength that satisfies both the Goldstein-Armijo conditions for $\|f_{PDD}\|^2$ and the restrictions $x + \alpha\Delta x > 0$ and $z + \alpha\Delta z > 0$.

The proof of the following theorem is similar to that of Theorem 6.1.

Theorem 7.1. *Let $\{v_i\}$ be the sequence generated by Newton's method applied to the equations (7.6). Then* $\lim_{i\to\infty} \|v_i - v^*(\mu)\| = 0$. ∎

At first sight, the Jacobian J_{PDD} looks more complicated than J_{PD}. However, the KKT matrix for the system is identical to that for the primal-dual method of Section 7.1. We have

$$\begin{pmatrix} ZX^{-1} & A^T \\ A & \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = - \begin{pmatrix} c - 2z + \dfrac{1}{\mu} Z^2 x - A^T y \\ Ax - b \end{pmatrix}. \qquad (7.7)$$

Hence, the search direction for the PD algorithm of Section 7.1 may be computed with little additional effort. A better direction can perhaps be constructed from the two search directions. The precise combination can be made dynamic and need be specified only after both directions are known.

The right-hand side of (7.7) is identical to the right-hand side for the KKT system of the dual algorithm. This implies that this algorithm is related to the dual barrier algorithm in the same way that the primal-dual algorithm of Section 7.1 is related to the primal. For example, a merit function based on the dual barrier function and dual constraint violations would enable the calculation of different steps in the primal and dual variables. In this case it is the step in the primal variables that may be chosen arbitrarily.

Note that any linear combination of the systems of equations (7.6), (7.1), (4.1) and (6.3) also leads to a similar algorithm. In particular, any linear combination that includes the primal-dual equations (7.6) (no matter how tiny a proportion) has the property that a suitable steplength exists for which x and z are positive.

8. The Treatment of Free Variables

If there are no bounds on a particular variable, no difficulty arises provided Δv is computed directly from the KKT system. However, a common approach for computing Δv is first to compute Δy using the Schur complement of the leading diagonal matrix. If the leading diagonal is singular, such an approach cannot be used. (We may always use the Schur complement of the nonsingular portion of the diagonal matrix but this is no longer a definite system.)

We shall consider just the primal algorithm, but the approach suggested here may be used in all the methods. For simplicity, assume that x_r is the only free variable. In place of $(\bar{f}_P)_r = z_r - \mu/x_r$ we now have $(\bar{f}_P)_r = z_r$, with $z_r^*(\mu) = 0$. As long as the explicit KKT system is solved, the effect of this equation on the Jacobian is inconsequential. However, in place of μX^{-2} we now have D, where $d_r = 0$ and $d_j = \mu x_j^{-2}$ for $j \neq r$. Hence D is singular and its Schur complement does not exist.

A means of circumventing this difficulty is to replace the equation $z_r = 0$ by an equation that ensures $z_r \to 0$ as $\mu \to 0$. For example, we could use

$$e^{x_r} z_r = \mu \qquad \text{or} \qquad z_r + \mu x_r = 0,$$

which give $d_r = z_r$ and $d_r = \mu$ respectively. In the first example, $z_r^*(\mu) = \mu e^{-x_r^*(\mu)}$ and we may keep $z_r \geq \sigma_Z > 0$. It follows that D is nonsingular and its Schur complement exists. In the second example, since z_r does not appear in the Jacobian, its sign or magnitude is not important. Likewise, the nonsingularity of J_P no longer depends on x_r, so there is no need to restrict the steplength for this variable.

9. Further Comments

A common practice in interior-point implementations is to define the steplength as some fixed percentage of the maximum feasible step. By contrast, all the algorithms described in this paper require some form of linesearch for the steplength. In practice this requirement has a minimal effect upon computation time, given the work needed to compute the search direction. Moreover, if η_1 is close to one and η_2 is close to zero, almost any step is likely to satisfy the Goldstein-Armijo conditions because all the linesearch functions are convex and increase rapidly near the boundary of the feasible region. In practice we have observed that the need to perform a linesearch arises only when there is significant numerical error in the search direction.

Currently the most efficient implementations use a predictor-corrector method to define the search direction (e.g. [LMS90,Meh90]). Such a strategy may be incorporated in the algorithms discussed here. The important point is to be able to fall back on a guaranteed method should the predictor-corrector direction fail to be a suitable descent direction for the merit function. A similar view was adopted by Mehrotra [Meh90].

It has not been our intent to compare the various algorithms in terms of performance. All the primal-dual algorithms have very similar theoretical properties, but only the primal-dual algorithm of Section 7.1 has been used in the principal known implementations [LMS89,LMS90,Meh90,GMPS91]. The key system of equations is "less nonlinear" than for the other three variations. Even so, in the neighborhood of the solution, the Jacobian behaves almost identically to the Jacobians of the other systems (as does the KKT matrix). It is not immediately apparent that this method is inherently superior to the others.

It may be that the best method is dependent on how the linear systems are solved. For example, all the methods may be implemented by solving systems of the form $ADA^T\Delta y = u$, where D is either X^2, Z^{-2} or XZ^{-1}. Suppose these systems are solved using a conjugate gradient method in which a preconditioner is based on periodically forming the Cholesky factors of ADA^T. The systems using $D = X^2$ should yield better preconditioners as the iterates converge because the ratio of consecutive values of any significant diagonal of D converges to one. When $D = XZ^{-1}$ or $D = Z^{-2}$, the significant diagonals correspond to the small elements of z. It is not obvious that the ratio of consecutive values of any such diagonal will behave as smoothly.

Our analysis is directed at the linear programming problem. However, extending the results to a smooth convex objective function is quite straightforward. The more challenging problem is to extend the results to nonconvex problems.

Acknowledgement

The authors are grateful to Joe Shinnerl for his careful reading of the manuscript and a number of helpful suggestions.

References

[FM68] A. V. Fiacco and G. P. McCormick (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York and Toronto.

[GMSTW86] P. E. Gill, W. Murray, M. A. Saunders, J. A. Tomlin and M. H. Wright (1986). On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method, *Mathematical Programming* 36, 183–209.

[GMPS91] P. E. Gill, W. Murray, D. B. Ponceleón and M. A. Saunders (1991). Solving reduced KKT systems in barrier methods for linear and quadratic programming, Report SOL 91-7, Department of Operations Research, Stanford University.

[KMY89] M. Kojima, S. Mizuno and A. Yoshise (1989). A primal-dual interior-point algorithm for linear programming, in N. Megiddo (ed.), *Progress in Mathematical Programming*, Springer-Verlag, New York, 29–47.

[KMM90] M. Kojima, N. Megiddo and S. Mizuno (1990). Theoretical convergence of large-step primal-dual interior-point algorithms for linear programming, Report RJ 7872 (72532), IBM Almaden Research Center, San Jose, CA.

[Lus88] I. J. Lustig (1988). Feasibility issues in an interior-point method for linear programming, Report SOR 88-9, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ.

[LMS89] I. J. Lustig, R. E. Marsten and D. F. Shanno (1989). Computational experience with a primal-dual interior-point method for linear programming, Report SOR 89-17, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ.

[LMS90] I. J. Lustig, R. E. Marsten and D. F. Shanno (1990). On implementing Mehrotra's predictor-corrector interior-point method for linear programming, Report SOR 90-3, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ.

[MMS89] K. A. McShane, C. L. Monma and D. F. Shanno (1989). An implementation of a primal-dual interior-point method for linear programming, *ORSA Journal on Computing* 1.

[Meg89] N. Megiddo (1986). Pathways to the optimal set in linear programming, in N. Megiddo (ed.), *Progress in Mathematical Programming*, Springer-Verlag, New York, 131–158.

[Meh90] S. Mehrotra (1990). On the implementation of a primal-dual interior-point method, Report 90-03, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

[MA89] R. D. C. Monteiro and I. Adler (1989). Interior path-following primal-dual algorithms, Part 1: Linear programming, *Mathematical Programming* 44, 27–41.

[OR70] J. M. Ortega and W. C. Rheinboldt (1970). *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, London and New York.

[Pon90] D. B. Ponceleón (1990). *Barrier methods for large-scale quadratic programming*, Ph.D. Thesis, Computer Science Department, Stanford University, Stanford.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1991	Technical Report

4. TITLE AND SUBTITLE
Primal-Dual Methods for Linear Programming

5. FUNDING NUMBERS
N00014-90-J-1242

6. AUTHOR(S)
Philip E. Gill, Walter Murray, Dulce B. Ponceleón
and Michael A. Saunders

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Operations Research - SOL
Stanford University
Stanford, CA 94305-4022

8. PERFORMING ORGANIZATION REPORT NUMBER
1111MA

9. SPONSORING MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research - Department of the Navy
300 W. Quincy Street
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
SOL 91-3

12a. DISTRIBUTION AVAILABILITY STATEMENT
UNLIMITED

12b DISTRIBUTION CODE
UL

13. ABSTRACT *(Maximum 200 words)*

Many interior-point methods for linear programming are based on the properties of the logarithmic barrier function. We first give a convergence proof for the (primal) projected Newton barrier method. We then analyze three types of barrier method that can be categorized as primal, dual and primal-dual. All three approaches may be derived from the application of Newton's method to different variants of the same system of nonlinear equations. A fourth variant of the same equations leads to a new primal-dual algorithm.

In each of the methods discussed, convergence is demonstrated without the need for a nondegeneracy assumption. In particular, convergence is established for a primal-dual algorithm that allows a different step in the primal and dual variables.

Finally, we describe a new method for treating free variables.

14. SUBJECT TERMS	15. NUMBER OF PAGES
linear programming; barrier methods, interior-point methods.	21 pages
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19 SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED			SAR